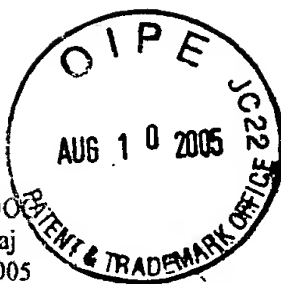


564227\_1.DOC  
MLW/AA/saj  
August 3, 2005



PATENT APPLICATION  
Attorney's Docket No.: 200304450-1 (0918.2033-000)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants: Simon Kasif, Beth T. Logan, Pedro J. Moreno and Baris E. Suzek

Application No.: 09/724,269 Group: 1631

Filed: November 28, 2000 Examiner: Mary K. Zeman

Confirmation No.: 7893

For: COMPUTER METHOD AND APPARATUS FOR UNIFORM  
REPRESENTATION OF GENOME SEQUENCES

CERTIFICATE OF MAILING	
I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as First Class Mail in an envelope addressed to Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450	
on <u>8/8/05</u>	<u>Sandra Jarmel</u>
Date	Signature
<u>Sandra Jarmel</u>	
Typed or printed name of person signing certificate	

DECLARATION OF BETH T. LOGAN AND PEDRO J. MORENO UNDER 37

C.F.R. 1.131

Commissioner of Patents and Trademarks  
P.O. Box 2327  
Alexandria, VA 22202

Sir:

We, Beth T. Logan and Pedro J. Moreno declare and state that:

1. We are the co-inventors of the above-identified application.

2. The invention described and claimed in the above-referenced application was conceived at least as early as May 2000, at least three months prior to August 22, 2000, which is the earliest priority date of the U.S. Pat. No. 6,834,239 to Lobanov *et al.*, cited by the Examiner as a basis of the rejection under 35 U.S.C. §102(e). The invention was reduced to practice by July 5, 2000. The conception, reduction to practice and diligence from conception to reduction to practice are evidenced by the attached Exhibits A through J which are described below in an account of the development of the invention to reduction to practice.

3. In late May 2000, Simon Kasif, Beth T. Logan, and Pedro J. Moreno, then all employees of COMPAQ Computer Corporation, subsequently acquired by Hewlett-Packard, and Baris E. Suzek, an intern at COMPAQ Computer Corporation during the Summer of 2000, convened to discuss a novel approach to protein classification whereby proteins would be represented by combining small sequences.

4. On June 9, 2000, Mr. Suzek recorded on the source controlled internal website, hereinafter, "the Website", that "we will try to find a novel approach to protein classification which will help biologists in finding: functional properties of proteins[,] structural properties of proteins[, and] evolutionary properties of proteins [...] ." Mr. Suzek also recorded that the project plan included "[d]evelop[ing] a tool to find the amino acid sequences (presumably short in length ) in the proteins that will help to classify them. Ideally, the tool will try to find the short sequences that best matches with the HMM [Hidden Markov Models] in a given database." A print-out of the Website as of June 9, 2000 is presented as Exhibit A. The relevant portions are highlighted.

5. Between June 9, 2000 and June 12, 2000 Mr. Kasif suggested studying the existing tools for sequence analysis and classification such as HMMER and BLIMPS. HMMER is a freely distributable software for protein sequence analysis using Hidden Markov Models (HMM), available from Washington University in St. Louis, Missouri, at the URL <http://hmmer.wustl.edu/>. BLIMPS (BLocks IMProved Searcher) is a searching

tool for BLOCKS database. The most current version of this freeware is available at the URL

<http://bioweb.pasteur.fr/seqanal/motif/blimps-uk.html>.

BLOCKS is a database of multiply aligned ungapped segments corresponding to the most highly conserved regions of proteins. It is maintained by Pittsburgh Supercomputing Center and is accessible through the URL

<http://www.psc.edu/general/software/packages/blocks/blocks.html>.

Mr. Kasif pointed out that it may be possible that the relationship of the segments of BLOCKS to the proteins could be analogous to the relationship of the phonemes to words. The implication of this suggestion is that protein sequences may be generated from the segments of BLOCKS. By June 12, 2000, Mr. Suzek recorded on the Website:

The consensus sequences of BLOCKS will be searched against PFAM to see if there is [sic] multiple hits per BLOCK [sic], which implies that BLOCKS can be building 'blocks' of domains: As a first step consensus seq[ue]nce[s] will be generated from BLOCKS database.

(PFAM, a Protein FAMilies database of alignments and HMMs, is a large collection of multiple sequence alignments and hidden Markov models covering many common protein domains and families. It is maintained by the Sanger Institute and is accessible through the URL <http://www.sanger.ac.uk/Software/Pfam/>.) A print-out of the Website as of June 12, 2000 is presented as Exhibit B. The relevant portions are highlighted.

6. During a meeting held on June 13, 2000, an idea of modeling each protein as a concatenation of BLOCKS segments was proposed. A concatenation of the BLOCKS segments led to the idea of converting each protein to a feature vector comprising information about the presence of each BLOCKS segment in a given protein sequence. On June 13, 2000, Mr. Suzek recorded on the Website:

Model proteins by concatenation of short "base units" separated by junk. This is similar to the PFAM domain idea except the base units are shorter than domains - more like the size of BLOCKS. Ideally model each base unit with a HMM.

On the same date, Mr. Suzek recorded on the Website under the heading Project Progress:

For each protein in the SCOP database, we will find the BLOCKS occurring in them. And generate a feature vector with the scores of BLOCKS found in them.

(The SCOP (Structural Classification of Proteins) database is a comprehensive ordering of all proteins of known structure, according to their evolutionary and structural relationships. SCOP is accessible at the URLs <http://scop.berkeley.edu/> or <http://scop.mrc-lmb.cam.ac.uk/scop/>.) A print-out of the Website as of June 13, 2000 is presented as Exhibit C. The relevant portions are highlighted.

7. By June 19, 2000, Mr. Suzek had generated feature vectors for all proteins in SCOP by scoring these proteins against the segments of the BLOCKS database (*i.e.* by counting the number of times each BLOCKS segments is contained in each SCOP protein). Mr. Suzek posted the generated vectors on the Website. A print-out of the Website as of June 20, 2000 is presented as Exhibit D. (See entry 7 under the heading Project Report. The relevant portions are highlighted.)

8. Following the generation of the feature vectors for a significant number of proteins, the question of classifying the proteins was addressed. On or before June 20, 2000, a brainstorming session was held during which various techniques for classifying multidimensional vectors were discussed. On June 20, 2000, Mr. Suzek recorded on the Website under the heading Brain Storming:

Given a feature vector whose entries are based on posterior probabilities of blocks, we could use SVD [Singular Value Decomposition] [...] to reduce the dimensionality of these huge vector (as many components as blocks!) and find the "important" components. Once this mapping from high dimension to low dimension is done we can also find natural clusters, use Gaussian modeling, classify etc.

Mr. Suzek further recorded on the Website that support vector machines (SVMs) can be used to classify protein families and that our results to the technique accepted in the art such as BLAST. (See Exhibit D, entries 8 and 9 under the heading Project Progress.)

9. On or before June 26, 2000, Mr. Suzek recorded on the Website the results of the comparison of the protein classification obtained using support vector machines to that

obtained using known methods described in Jaakkola *et al.* "A Discriminative Framework for Detecting Remote Protein Homologies", J. Comp. Biol., Vol. 7, Num. 1/2 (2000). A print-out of the Website as of June 26, 2000 is presented as Exhibit E. See entry 9 under the heading Project Progress. The relevant portions are highlighted. (The abbreviation "FPR" stands for False Positive Rate.)

10. At a meeting that took place on or around June 27, 2000, methods of scoring proteins that contained a given segment of the BLOCKS database more than once were discussed. Two approaches were proposed. The first approach was to add the scores and second approach was to take the maximum of the scores. On or before July 7, 2000, Mr. Suzek made a corresponding entry on the Website. A print-out of the Website as of July 7, 2000 is presented as Exhibit F. See entry 7 under the heading Project Progress. The relevant portions are highlighted.

11. By July 5, 2000, Mr. Suzek reported completion of training the support vector machines embodiment of the invention based on protein classification obtained using methods described in Jaakkola *et al.* "A Discriminative Framework for Detecting Remote Protein Homologies", J. Comp. Biol., Vol. 7, Num. 1/2 (2000). See entry 9 under the heading Project Progress of Exhibit F. The relevant portions are highlighted.

12. In early August 2000, Mr. Suzek prepared a presentation based on the results of his summer internship at Compaq Computer Corp. A print out of this presentation in the PowerPoint format is presented as Exhibit G. Mr. Suzek's presentation was copied to the Website on August 11, 2000.

13. By August 3, 2000, we commenced drafting an invention disclosure and on August 21, 2000, the final version of the invention disclosure was sent to Compaq legal counsel. A copy of an email, with attachment, to Mr. R. Reed, an engineering liaison to a legal counsel for Compaq Computer Corp., is presented as Exhibit H. In section 4 of the invention disclosure for (Exhibit H), we report implementation of the invention in software between June 15 and July 31, 2000.

14. On September 15, 2000, Mr. R. Lange, a Legal Counsel for Compaq Computer Corp., contacted Ms. MaryLou Wakimura, a principal at the law firm of Hamilton, Brook, Smith & Reynolds with a request to prepare the patent application based on the research work described above. A copy of the email to Ms. Wakimura is presented as Exhibit I.

15. On October 5, 2000 we met with Ms. Wakimura to discuss drafting the patent application. A copy of the email from Ms. Logan to Ms. Wakimura scheduling this meeting is presented as Exhibit J.

16. Through October and November 2000, Ms. Wakimura produced a patent application which was filed in the USPTO on November 11, 2000 as evidenced by the present subject patent application.

17. I hereby acknowledge that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

  
\_\_\_\_\_  
BETH T. LOGAN

AUGUST 3 2005  
\_\_\_\_\_  
Date

\_\_\_\_\_  
PEDRO J. MORENO

\_\_\_\_\_  
Date

### **Index of Exhibits**

Exhibit A: A print-out of the source-controlled internal Compaq Computer Corp. website as of June 9, 2000.

Exhibit B: A print-out of the source-controlled internal Compaq Computer Corp. website as of June 12, 2000.

Exhibit C: A print-out of the source-controlled internal Compaq Computer Corp. website as of June 13, 2000.

Exhibit D: A print-out of the source-controlled internal Compaq Computer Corp. website as of June 20, 2000.

Exhibit E: A print-out of the source-controlled internal Compaq Computer Corp. website as of June 26, 2000.

Exhibit F: A print-out of the source-controlled internal Compaq Computer Corp. website as of July 7, 2000.

Exhibit G: A print out of the presentation slides by Mr. Suzek.

Exhibit H: A copy of an email dated August 21, 2000, from Ms. Logan to Mr. Reed and the attached invention disclosure.

Exhibit I: A copy of an email dated September 15, 2000, from Ms. Wakimura to Mr. Lange regarding preparation of the subject patent application.

Exhibit J: A copy of an email dated September 22, 2000, from Ms. Logan to Ms. Wakimura regarding scheduling a meeting to discuss the subject patent application.